

## **AN ECONOMETRICAL MODEL FOR CALCULATING THE ROMANIAN GROSS DOMESTIC PRODUCT**

*Ana Maria Mihaela Iordache*<sup>1</sup>

*Ionela Catalina Tudorache*<sup>2</sup>

*Mihai Tiberiu Iordache*<sup>3</sup>

### **Abstract**

*This article analyzes the possibility of determining the Romanian Gross Domestic Product on the basis of a linear model, based on macroeconomic indicators such as unemployment, inflation, exchange rate. The proposed solution allows forecasting the Gross Domestic Product for the following period of time based on equation determined from the model in order to develop scenarios for improving the macroeconomic situation of Romania. The article also can be a starting point in the field of strategic investments, both at macro and micro level. Any foreign investor, before starting a business in another country, calculated and also made predictions about the economic situation of that country, in order not to lose the investment made.*

**Key words:** regression, GDP, inflation, unemployment, forecast

### **Presenting the model used in application**

The main purposes of this model are to make forecasts, that are mathematically correct, as well as to verify the economic theory concerning the influence that different rates might have on the Gross Domestic Product.

We chose four different time series for a period of 13 years (1998-2010) [6]:

- the unemployment rate, calculated as the rate between the number of unemployed persons divided to the number of all individuals in the labor force;
- the annually inflation rate;
- the exchange rate euro/national currency (Romanian ron);
- the GDP at market prices, expressed in millions of national currency.

The regression model describes the relation (linear, hyperbolic, exponentially, or logarithmic) between a variable/variables that is/are the cause/causes, represented on the Ox axis, and a variable that is the effect, represented on the Oy axis.

The model presented below is used to determine a relation cause/causes-effect, to predict a dependent variable when the independent variables (one, two or many) are given and to explain the effect when we know the cause/causes.

The general linear model is presented by the equation:

$$y = a + bx_1 + cx_2 + dx_3 + \varepsilon$$

Where:

---

<sup>1</sup> Assitant Lecturer, School of Computer Science for Business Management, Romanian-American University

<sup>2</sup> Student, Romanian-American University

<sup>3</sup> Engineer, IMSAT Bucharest

- the parameters  $b, c, d$  show the proportional modification of the effect variable ( $Y$ ) when the cause variables ( $X_1, X_2, X_3$ ) modify, in average, with one unit;
- parameter  $a$  indicates the intersection point between the graph and the Oy axis;
- $\varepsilon$  is the random error and it represents the residual component of the model for each unit. That means that it shows which part of the  $Y$  value cannot be measured by the model, by the relation with the  $X$  variable.

In our case, we will consider the regression model represented by equation:

$y = \alpha + \beta x_1 + \gamma x_2 + \delta x_3 + \varepsilon$ , and the regression equation as  $\hat{y} = a + bx_1 + cx_2 + dx_3$ , where  $\hat{y}$  is the adjusted  $y$ , the line without the influence of  $\varepsilon$ . [1]

The independent variables from the model have the following meaning, from the economic point of view:

- $X_1$  represents the annually unemployment rate, expressed in percent;
  - $X_2$  is the annually inflation rate, also expressed in percent;
  - $X_3$  is the annually exchange rate between euro and ron (The euro currency was introduced at 1<sup>st</sup> January 1999, so for 1998, we took the exchange rate between ECU - European Currency Unit - and Romanian ron);
- $Y$  is the GDP at market prices, expressed in millions of national currency (mil. ron).

### Calculation and interpretation the coefficients of the model

For calculating the four coefficients from the model, we used the E-views software [3], which generates different types of outputs, for variables ( $X_1, X_2, X_3$  and  $Y$ ) imported from a document created in Excel software (see figure 1).

	A	B	C	D	E
1		GDP (mil. lei - market prices)	Inflation rate	Unemployment rate	Exchange rate euro-leu
2	1998	37055.1	59.1	6.3	0.9985
3	1999	55191.4	45.8	7.1	1.6345
4	2000	80984.6	45.7	7.3	1.9922
5	2001	117945.8	34.5	6.8	2.6004
6	2002	152017.0	22.5	8.6	3.1270
7	2003	197427.6	15.3	7	3.7551
8	2004	247368.0	11.9	8.1	4.0510
9	2005	288954.6	9.1	7.2	3.6209
10	2006	344650.6	6.6	7.3	3.5258
11	2007	416006.8	4.9	6.4	3.3353
12	2008	514700.0	7.9	5.8	3.6826
13	2009	498007.5	5.6	6.9	4.2399
14	2010	513640.8	6.1	7.3	4.2122
15					

Figure 1. Excel preview - the time series used in application [9]

Equation: OUTPUT_COEF    Workfile: EViews WORK				
View   Procs   Objects   Print   Name   Freeze   Estimate   Forecast   Stats   Resids				
Dependent Variable: Y				
Method: Least Squares				
Date: 07/22/11   Time: 12:42				
Sample: 1 13				
Included observations: 13				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	837618.5	321338.5	2.606655	0.0284
X1	-88200.93	33991.60	-2.594786	0.0290
X2	-5558.213	4258.796	-1.305114	0.2242
X3	54609.50	78781.68	0.693175	0.5057
R-squared	0.857226	Mean dependent var	266457.7	
Adjusted R-squared	0.809634	S.D. dependent var	177309.9	
S.E. of regression	77361.94	Akaike info criterion	25.59804	
Sum squared resid	5.39E+10	Schwarz criterion	25.77187	
Log likelihood	-162.3872	F-statistic	18.01218	
Durbin-Watson stat	1.350797	Prob(F-statistic)	0.000382	

Figure 2. E-views output - the estimate equation

Figure 2 represents the first output from E-views which it provides the coefficients for the regression model, the standard errors for the dependent variables and their probabilities [8].

Therefore, the model is represented by the following equation:  $\hat{y}_t = 837618.5 - 88200.93x_{1t} - 5558.213x_{2t} + 54609.50x_{3t}$  where  $\hat{y}_t$  means the GDP in year t.

- $a=837618.5$  represents the intersection point between the regression line and Oy axis. This value also shows the influence of other factors, that we don't take into account, on the GDP. The National Institute of Statistics calculates the GDP using three different methods with the same result, but, in spite of this, the result is an estimated value, due to many other factors that are immeasurable, or undetectable, such as the underground economy. This means that when  $x_1=x_2=x_3=0$  then  $y=837618.5$ . But, in terms of real economy,  $x_3$  will never reach the zero level, because it is the exchange rate euro-ron. Also, it could be equal with 1 when the parity purchase power of ron will be equal to the euro's one. In this case, the GDP will be 892228 mil ron.

- $b = - 88200.93$ , and  $c = - 5558.213$ , we have the situation: each time the unemployment rate and the inflation rate increase with one unit the GDP decreases, in average, with 88200.93 and 5558.213 million units. The value of  $d$  variable is 54609.50, it means that each time the exchange rate euro-ron increases with one unit, the GDP increases, in average, with 54609.50 million units.

### Testing the multifactorial regression model and the parameters used

If we want to see if the parameters  $a$ ,  $b$ ,  $c$  and  $d$  are correct, and they are not affected by errors, then we must use the Student statistic [5].

For testing  $\alpha$  parameter, we consider the following hypotheses:

$$H_0: \alpha = 0$$

$$H_1: \alpha \neq 0$$

The standard deviation of a variable is  $s_a = 321338.5$  (standard error).

The value of the statistic test is  $t = \frac{a-\alpha}{s_a} = \frac{837618.5-0}{321338.5} = 2.60665$

Because  $t = 2.60665$  with a probability (p-value) of 0.0284, we consider the null hypothesis false, and that a parameter is significantly from statistical point of view, which means that it is a valid parameter.

For testing the  $\beta$  parameter, we consider the hypotheses:

$$H_0: \beta = 0$$

$$H_1: \beta \neq 0$$

The standard deviation of  $b$  is  $s_b = 33991.60$  (standard error).

The value of the statistic test is  $t = \frac{b-\beta}{s_b} = \frac{-88200.93-0}{33991.60} = -2.59478$

Because  $t = -2.59478$  with a probability (p-value) of 0.0290, we consider false the null hypothesis, at a significance limit level of 5%, which means that the unemployment rate has a significantly influence on GDP.

For  $\gamma$  parameter, we considered the hypotheses:

$$H_0: \gamma = 0$$

$$H_1: \gamma \neq 0$$

The standard deviation of  $c$  is  $s_c = 4258.796$  (standard error).

The value of the statistic test is  $t = \frac{c-\gamma}{s_c} = \frac{-5558.213-0}{4258.796} = -1.30511$

Because the variable  $t$  is equal with -1.30511, with a probability (p-value) of 0.2242, we consider correct (true) the null hypothesis, at a significance limit level of 5%, which it means that the inflation rate has not a significantly influence on GDP. This parameter has not a good influence on the model.

For  $\delta$  parameter, we consider the following hypotheses:

$$H_0: \delta = 0$$

$$H_1: \delta \neq 0$$

The standard deviation of  $d$  is  $s_d = 78781.68$  (standard error).

The value of the statistic test is  $t = \frac{d-\delta}{s_d} = \frac{54609.50-0}{78781.68} = 0.69317$

Because the variable  $t$  is equal with 0.69317, with a probability (p-value) of 0.5057, we consider true the null hypothesis, at a significance limit level of 5%, which it means that the exchange rate euro-ron has not a significantly influence on GDP. Also, this parameter has not a good influence on the model.

The conclusion of these four testes is that only the intercept parameter ( $a$ ) and the unemployment rate parameter were accepted as mostly influencing the GDP. Even this, it is too early to affirm that the model is not valid.

The regression model can be testing using the Fisher test (or Test F). In this way, there are two hypotheses:

$H_0$  : the model is correct (valid)

$H_1$  : the model is incorrect (invalid)

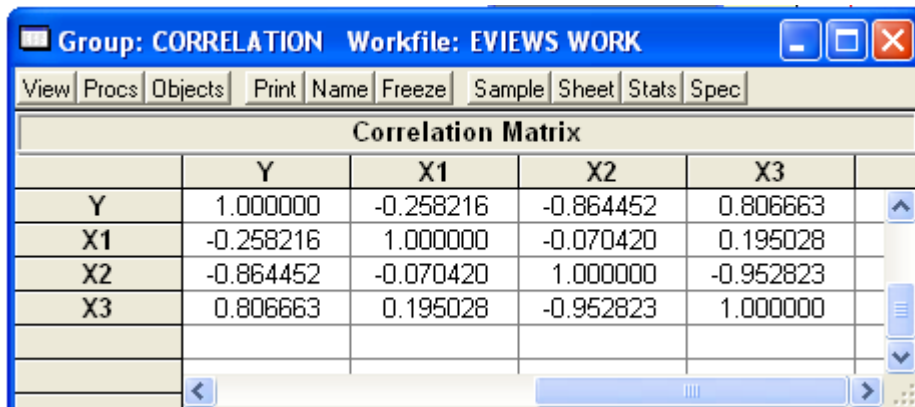
The significance testing can be made using the F test, knowing the two values:  $F_{calculated}$  and  $F_{\alpha, v_1, v_2}$ , the theoretical value of variable “F”, which is given in the Fisher-Snedecor repartition table.  $F_{\alpha, v_1, v_2}$  represents the value of the indicator at a significance level  $\alpha$ , with number of degrees of freedom  $v_1=k$ , and  $v_2=n-k-1$ .

In general, the decision rule are represented by two cases: we accept  $H_0$  and deny  $H_1$  if  $F_{calculated} < F_{\alpha, v_1, v_2}$ , or we deny  $H_0$  and accept  $H_1$  if  $F_{calculated} > F_{\alpha, v_1, v_2}$ .

In this case, from figure 2:  $F_{calculated} = 18.01218 > F_{0.05; 3, 9} = 5.7394$ , we consider false the null hypothesis and true the opposite, and the variable Prob (F-statistic) is very small (0.000382), which means that, even though only two parameters from four were accepted as having a value different of zero, the entire model is valid.

### The correlation matrix and the errors of the model

- The correlation matrix (figure 3) shows the coefficients of correlation between variables that have values between -1 and 1 and it shows how much from  $Y$  variation is explained by  $X$  variation ( $X \in \{X_1, X_2, X_3\}$ ) [4].
- In general, if the coefficient of correlation between  $X$  and  $Y$  is close to zero (positive or negative), the influence of  $X$  against  $Y$  is very low. On the other side, if it is closer to -1, the influence is strong, but negative, and closed to 1, it means that the influence is strong and positive [10].
- 



- Figure 3.Eviews output - The correlation matrix

In figure 3 it is shown all three correlation coefficients:

- $r_{X_1Y} = -0.258216$  means that the annually unemployment rate has a negative weak (low) influence on the annually GDP;
- $r_{X_2Y} = -0.864452$ , it means that the inflation rate has a negative strong (high) influence on GDP;
- $r_{X_3Y} = 0.806663$ , it means that the annually exchange rate euro-ron has a positive strong influence on GDP.

So far, we considered that the model is not affected of errors very much. Any regression model requires testing all fundamental hypotheses, the Gauss-Markov hypotheses, in order to consider the model as being correct from econometrical point of view.

The most important two hypotheses refer to  $\varepsilon$  variable are:

1. The White heteroskedasticity test (figure 4) shows that the mean of  $\varepsilon$  variable different of zero, and it's variance is not constant.

#### White Heteroskedasticity Test:

F-statistic	1.094931	Probability	0.457576
Obs*R-squared	6.794544	Probability	0.340266

#### Test Equation:

Dependent Variable: RESID^2

Method: Least Squares

Date: 12/10/11 Time: 20:09

Sample: 1 13

Included observations: 13

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.35E+11	1.52E+11	-0.888634	0.4084
X1	3.80E+10	3.70E+10	1.028041	0.3436
X1^2	-2.86E+09	2.56E+09	-1.116720	0.3068
X2	1.01E+09	9.20E+08	1.096716	0.3148
X2^2	-13371705	25593352	-0.522468	0.6201
X3	-8.16E+09	4.68E+10	-0.174226	0.8674
X3^2	2.72E+09	6.30E+09	0.431489	0.6812
R-squared	0.522657	Mean dependent var	4.14E+09	
Adjusted R-squared	0.045315	S.D. dependent var	6.40E+09	
S.E. of regression	6.25E+09	Akaike info criterion	48.25463	
Sum squared resid	2.35E+20	Schwarz criterion	48.55883	
Log likelihood	-306.6551	F-statistic	1.094931	
Durbin-Watson stat	2.272822	Prob(F-statistic)	0.457576	

Figure 4. The White Heteroskedasticity Test

By calculating an auxiliary regression using the errors squares, the variable  $a$  remains constant, the explicative variables of the original model, their squares and crossed math multiplication, we retain the coefficient of multiple determination  $R_a^2$ .

White showed that the statistic  $W=n*R_a^2$  follows a chi-squared distribution, with the number of degree of freedom equal to the number of regression extern variables from the auxiliary model, there are six in our case.

In figure 4,  $W= 6.794544$ , with a probability of 0.340266.

$H_0$ : homoskedasticity exists in the model

$H_1$ : heteroskedasticity exists in the model

Because  $W_{\text{calculated}} = 6.79 > W_{\text{critic}} = 1.635$ , and, moreover  $P(W) = 0.34 > 0.05$ , we consider false the null hypotheses, which it means that the errors are heteroskedastic (the mean is not zero, and the variance is not constant).

2. The independence, the absence of autocorrelation, of  $\varepsilon$  variable values, which gives us information about the intensity of connection between  $y_t$  and  $y_{t-k}$  (if it exists!).

Here, we can use the Durbin-Watson statistic (see figure 2). This test was elaborated in 1950, by Durbin J. and Watson G.S. SO, from figure 2, the Durbin-Watson statistic is  $DW = 1.350797$ . From the distribution table of this statistic, we take two values:  $d_1 = 0.72$  and  $d_2 = 1.82$ .

Because  $d_1 < DW < 4 - d_2$ , we can affirm, with a probability of 95%, that the autocorrelation of errors is absent.

### **The forecast of GDP based on the model**

As we shown previously, the regression model studied is valid with a high probability and, therefore, it allows some different types of forecasting:

1. For 2011, the inflation rate is forecasted by National Romanian Bank at 5.1%, the unemployment rate will be 8%, while the exchange rate euro-ron is estimated at 4.1403. Based on this dates, it would be very interesting to know what would be the value of GDP in 2011.
  - It is known that:  $\hat{y}_t = 837618.5 - 88200.93x_{1t} - 5558.213x_{2t} + 54609.50x_{3t}$ , where  $\hat{y}_t$  is the GDP in year t.
  - For 2011,  $\widehat{y}_{2011} = 837618.5 - 88200.93x_{1,2011} - 5558.213x_{2,2011} + 54609.50x_{3,2011}$  where  $x_{1,2011}$ ,  $x_{2,2011}$  and  $x_{3,2011}$  are known. So, in 2011, the GDP will be 329764 million ron.
2. In another case, if the estimated GDP in 2011 is 544352.2, the inflation rate is 5.1%, while the exchange rate euro-ron is estimated at 4.1403, it would be very interesting to know what would be the unemployment rate from Romania, in 2011.
  - Based on the same equation, the unemployment rate for 2011 would be 5.56%.

### **Conclusion**

Obviously, the predictions made, based on this regression model, can be strongly affected by errors, which it means that the first hypothesis is not accomplished. To reduce or to avoid the heteroskedasticity of the errors and to improve the model, it is highly recommended, in case we know the variances of errors, to transform the model, either by using the radical model, by dividing each term with  $\sqrt{X_i}$ , either by using the natural logarithm,  $\ln y = a + b \ln x_1 + c \ln x_2 + d \ln x_3 + \varepsilon$ . [5]

The model presented in the article includes only three parameters: the unemployment rate, the inflation rate and the exchange rate euro-ron. Estimating the GDP or another independent variable involves taking into account a big amount of information: data series that contains thousands or millions of values and, of course, hundreds of independent variables. What we

tried to prove is that it is possible to create a mathematical model, at a large scale, from a small sample, in order to test the economic theory based on relations.

## **Bibliography**

- [1] Anghelache Constantin, Tratat de statistică teoretică și economică, Editura Economică, București, 2008, ISBN 9789737093806
- [2] David Cristian, Voineagu Vergil, Analiza statistică a eforturilor și efectelor aderării României la Uniunea Europeană, București, editura ASE, 2008
- [3] How to Use EViews (Econometric Views)-<http://mlovell.web.wesleyan.edu/e280/EviewsHow.PDF>
- [4] Iacob, A. I, Tanasoiu, O., Econometrie: studii de caz, Ed. ASE, București, 2005
- [5] The Econometrics Journal-<http://www.res.org.uk/econometrics/econometricshome.asp>
- [6] Văcărel Iulian, Bistriceanu Gh, Bercea Florian, Anghelache Gabriela, Bodnar Maria, Moșteanu Tatiana, Georgescu Florin, Finanțe publice, ediția a IV-a, Editura Didactică și pedagogică, București, 2004, ISBN: 973-30-2635-2
- [7] [http://www.acc.ncku.edu.tw/chinese/faculty/wangt/About%20Class/Class/eviews\\_handout.pdf](http://www.acc.ncku.edu.tw/chinese/faculty/wangt/About%20Class/Class/eviews_handout.pdf)
- [8] <http://www.econ.kuleuven.be/public/ndbae76/econometrics/eview4.0.pdf>
- [9] <http://epp.eurostat.ec.europa.eu>
- [10] [http://en.wikibooks.org/wiki/Econometric\\_Theory](http://en.wikibooks.org/wiki/Econometric_Theory)



